# What we estimate when we estimate dynamic causal effects in panel data

February 10, 2026

**Abstract**

Policy-relevant decisions depend on causal effects in the current or next period (e.g., FATT, FATU), but standard panel-data methods target backward-looking estimands such as the average treatment effect on the treated (ATT) or group-time ATTs over the observed window. We formalize this gap using future causal effects and show that identifying them requires either taking dynamic effects *less* seriously—by assuming time homogeneity so that backward-looking ATT or group-time effects equal the FATT—or *more* seriously—by specifying how effects evolve and extrapolating. Under time homogeneity, the FATT is identified by the ATT or by a convex combination of within-group ATTs; under a parametric model for group-time effects, the FATT is identified by extrapolating the fitted model to the future period. We give efficient influence functions for the FATT under both paths so that standard errors and confidence intervals propagate correctly from first-stage estimators.

## 1 Introduction

There is a fundamental and unacknowledged tension at the heart of policy analysis and causal inference with panel data more broadly [Egami and Hartman, 2023, Heckman and Vytlacil, 2001]. This tension arises because of the possibility of dynamic causal effects (effects that change over time), which, as we'll argue below, preclude an important type of external validity: if policy effects are truly unrestrictedly dynamic, then it is impossible to use historical data to inform *any* current decisions, which decisions are presumably the reason we care about estimating any causal effects in the first place.

Most causal inference rests on an implicit assumption of time homogeneity. We might call this assumption the Fundamental Promise of Causal Inference: causal effects estimated (by necessity) on data collected in the past can be used to inform decision-making in the present [Egami and Hartman, 2023, Khosrowi, 2019, Cartwright, 2007]. If two identical clinical trials – same target population and inclusion criteria, same treatment, same outcome – were expected to not only give different results but in fact target different causal effects simply because of the passage of time, that would be a major crisis for causal inference. A durable understanding of causal effects would be elusive.

Public policies are widely understood to violate this assumption and are assumed to have effects that inherently change over time [Lucas Jr, 1976, Ericsson et al., 1998, Estrella and Fuhrer, 1999]. Policies are introduced into dynamic contexts, may be administered or enforced differently over time, and may themselves cause changes in public behavior and attitude. Technological advancement, public opinion, economic conditions, and new research may all modify the causal effects policies.

This basic assumption of dynamic effects has driven a great deal of recent advances in methods for causal inference with panel data [Sun and Abraham, 2021, Wang et al., 2024]. Recent surveys clarify how heterogeneity and dynamics enter identification and estimation in this setting [Arkhangelsky and Imbens, 2024, De Chaisemartin and d'Haultfoeuille, 2023]. It is now standard to assume that causal effects vary both with time of treatment and time since implementation. These effects (often called "group-time" effects) may be of interest in themselves but more typically are aggregated into coarser effects like an "overall" average treatment effect on the treated (ATT). Assuming such dynamic effects solves a problem with classical methods like two-way fixed effects: they may incur bias in the presence of time-heterogeneous effects because they implicitly include comparisons between units treated at different times in their estimate of the ATT [Sun and Abraham, 2021].

While methods assuming dynamic effects have been developed for good reasons, they strike at the heart of the Fundamental Promise of Causal Inference. In the presence of dynamic effects, it is not clear how one can connect causal effects estimated on historical data to any decision one would make now or in the future [Egami and Hartman, 2023, Devaux and Egami, 2022, Khosrowi, 2019]. The question of when and how historical evidence can inform current or future decisions is central to the literature on generalizing and transporting causal effects [Cole and Stuart, 2010, Tipton, 2013, Manski, 2013]. If one estimates the effect of a public policy to be beneficial (e.g., improves life expectancy) among states that implemented it in historical data, does that mean that it would be wise to maintain that policy in places that have already implemented and/or to implement it in places it has not been implemented? If effects are unrestrictedly dynamic, it's impossible to say.

We seek to resolve this tension between dynamic effects and the Fundamental Promise by encouraging researchers to either take the possibility of dynamic effects *less* seriously or *more* seriously. Taking it less seriously involves assuming some level of time homogeneity in causal effects. This would allow causal effects that would inform current and future decisions to be identified in historical data. We argue that an informal version of this is currently standard practice, where average treatment effects estimated on historical data are treated as if they are stable enough over time to inform current debates [Gabler et al., 2009, Cintron et al., 2022]. The notion of policy-relevant treatment effects formalizes which estimands answer counterfactual policy questions [Heckman and Vytlacil, 2001, Sasaki and Ura, 2023]. On the other hand, taking dynamic effects more seriously would require formally modeling how dynamic effects change over time and using such a model to extrapolate into the future [Khosrowi, 2019, Maeba, 2024], as in structural approaches that enable extrapolation across regimes [Heckman and Vytlacil, 2005, Pearl and Bareinboim, 2022].

We introduce new causal estimands that formalize the notion that causal inference is inherently future-oriented, which we call "future causal effects." We locate the estimands outside the time range of the current study and show how each of these paths – the more serious and less serious ones – may be used to identify the future causal effects and under what type of

assumptions. Across fields, three broad strategies emerge for making causal claims about the future: structural or parameter invariance (e.g., policy-invariant deep parameters); statistical transport or reweighting of effects across populations; and explicit modeling of heterogeneity or dynamics. Our future-looking estimands sit at the intersection of these traditions.

## 1.1 Related work

Recent advances in causal inference for panel data and difference-in-differences provide the backbone for our setting. Surveys by Arkhangelsky and Imbens [2024] and De Chaisemartin and d'Haultfoeuille [2023] clarify how heterogeneity and dynamics enter identification and estimation; the group-time and staggered-adoption literature [Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Roth and Sant'Anna, 2023, Baker et al., 2025] addresses inference for backward-looking effects when treatment timing varies. Work on testing parallel trends and time homogeneity—including the limitations of conventional pre-trend tests [Freyaldenhoven et al., 2019], the pitfalls of conditioning on passing such tests [Roth, 2022], inference under bounded trend violations [Rambachan and Roth, 2023], and equivalence tests for pre-trends [Dette and Schumann, 2024]—connects directly to the testable implications of our time-homogeneity assumptions. These lines of work focus on identification and inference for effects in the observed window; we instead target estimands that are relevant for current or future policy decisions.

The disconnect between typical estimands (ATT, group-time effects) and policy-relevant or future causal questions has been formalized in several ways. Heckman and Vytlacil [2001] define policy-relevant treatment effects for counterfactual policy changes; Sasaki and Ura [2023] provide estimation and inference for such parameters. Forastiere et al. [2025] formalize forecasting the causal effect of *re-implementing* an intervention at a future time; we focus on the effect of the policy in the current or next period ($p + 1$)—continuing or repealing for current adopters, adopting or not for non-adopters—and potentially beyond. Gische et al. [2021] distinguish forecasting the causal effect of an intervention from predicting future outcomes; we similarly distinguish backward-looking estimands (e.g., the ATT in the observed window) from forward-looking, policy-relevant estimands such as the FATT and FATU at time $p + 1$.

External validity and transportability are central to our concern that backward-looking estimands may not inform current decisions. Egami and Hartman [2023] provide a framework for design and analysis; Pearl and Bareinboim [2022] and Bareinboim and Pearl [2012] give graphical and identification results for transporting causal effects. The literature on generalizing from randomized trials to target populations [Cole and Stuart, 2010, Tipton, 2013] and on robustness to external validity bias [Devaux and Egami, 2022] is closely related. Khosrowi [2019] and Khosrowi [2023] examine the assumptions and uncertainties underlying extrapolation; Manski [2013] advocates bounded identification when extrapolation is weak. Shpitser [2013] extend transportability to longitudinal settings.

The Lucas Critique and the question of policy invariance underlie our discussion of when effects can be taken as stable over time. Lucas Jr [1976] argues that policy evaluation requires parameters that are invariant to the policy regime; Ericsson

et al. [1998] and Estrella and Fuhrer [1999] formalize and test such invariance. Structural and parametric approaches to extrapolation—including Heckman and Vytlacil [2005] on policy-invariant parameters, Pearl and Bareinboim [2022] on transportability, and Brodersen et al. [2015] on Bayesian structural time series—show how functional assumptions enable extrapolation across regimes or time periods.

Philosophical work on causal invariance and extrapolation [Cartwright, 2007] and empirical evidence on heterogeneous treatment effects in policy and program evaluation [Gabler et al., 2009, Cintron et al., 2022] further motivate the need to be explicit about which effects are policy-relevant and under what assumptions they can be identified from historical data.

Our future-looking estimands sit at the intersection of three strategies for making causal claims about the future: structural or parameter invariance, statistical transport or reweighting, and explicit modeling of heterogeneity or dynamics.

## 2 Setting and context

Suppose we measure an indicator of a policy $A_{it}$, a possibly affected outcome $Y_{it}$, and time-invariant covariates $\mathbf{X}_i$ for $i = 1, ..., n$ units (often the units are geographic entities, like states, counties, etc.) and $t = 1, ..., p$ timepoints. For ease of exposition, we take $A_{it}$ to be a policy, but all that follows applies equally to other types of interventions. Define the counterfactual value that $Y_{it}$ would take if the policy were set, possibly contrary to fact, to $A_{it} = a$ as $Y_{it}(a)$. We assume no interference: each unit's potential outcomes depend only on that unit's treatment, so $Y_{it}(a)$ is well-defined. Let $\mathcal{O} = \{\mathbf{X}_i, \mathbf{Y}_i(0), \mathbf{Y}_i(1), \mathbf{A}_i\}_{i=1,...,n} \sim \mathbb{P}$ where $\mathbf{Y}_i(a) = \{\mathbf{Y}_{it}(a)\}_{t=1,...,p}$ and $\mathbf{A}_i = (A_{it})_{t=1,...,p}$. We extend the setting to future time periods when defining future estimands below. The distribution of data at time $t$ is $\mathbb{P}_t$. For integer $k \geq 1$, we write $\mathbb{P}_{p+k}$ for the distribution of $(\mathbf{X}_i, Y_{i,p+k}(0), Y_{i,p+k}(1), A_{i,p+k})$ in the same superpopulation $k$ periods ahead (i.e., the marginal at $t = p + k$ of the joint distribution governing the process through $p + k$); $\mathbb{P}_{p+1}$ is the case $k = 1$.

It is typical (though not necessary) in such studies to take the target of estimation to be the average treatment effect on the treated (ATT). The ATT may be defined as

$$\theta = \mathbb{E}_{\mathbb{P}}\{Y_{it}(1) - Y_{it}(0)|A_{it} = 1\}.$$

One virtue of the ATT is that $Y_{it}(1)$ is always observed under the conditioning event $A_{it} = 1$, so identification of the ATT is facilitated by only requiring causal inference machinery to estimate $\mathbb{E}_{\mathbb{P}}\{Y_{it}(0)|A_{it} = 1\}$.

In the presence of heterogeneity – especially heterogeneity in how policy effects change over time – there have been proposals to instead decompose the ATT into separate sub-ATTs, with an example being referred to as group-time ATTs [Sun and Abraham, 2021].

$$\theta_{gt} = \mathbb{E}_{\mathbb{P}_t}\{Y_{it}(1) - Y_{it}(0)|A_{it} = 1, G_i = g\}$$

where $G_i = \min_t tI\{A_{it} = 1\}$, the time when the policy first occurred, and $\mathbb{P}_t$ is the distribution of the data at time $t$: $\mathcal{O}_t = \{\mathbf{X}_i, Y_{it}(0), Y_{it}(1), A_{it}\}_{i=1,...,n} \sim \mathbb{P}_t$. This is important to solve some infelicities with trying to estimate $\theta$ directly with a single linear model [Sun and Abraham, 2021] and to draw attention to the fact that causal effects may both vary across calendar time and have different effects as implementation continues – i.e., effects may differ in the first year of implementation from the fifth year, say. These sub-ATTs are typically then aggregated up into an overall effect estimate across groups/times.

## 2.1 Identification of the backward-looking ATT

The results in Sections 4 and 5 assume that the researcher has secured identification of the ATT $\theta$ and, where relevant, the group-time ATTs $\theta_{gt}$, under some design. For example, under parallel trends in difference-in-differences, $\theta_{gt}$ is identified from the observed data [Callaway and Sant'Anna, 2021, Sun and Abraham, 2021]; under unconfoundedness or other designs, different identification formulas apply. We do not restrict the first-stage identification strategy: it may be difference-in-differences, synthetic control, or another approach. The contribution of this paper is to show when and how these backward-looking estimands coincide with or can be used to identify the policy-relevant FATT (and related future estimands), given either time homogeneity assumptions (Section 4) or a parametric model for dynamics (Section 5). Thus the full identification argument has two steps: (a) under the researcher's design, $\theta$ or $\theta_{gt}$ are identified from observables; (b) under the extrapolation assumptions we state below, the FATT is identified from $\theta$ or $\theta_{gt}$.

However, regardless of how they are aggregated, these effects are typically not directly relevant to current policy or decisionmaking [Egami and Hartman, 2023, Khosrowi, 2019]. These ATTs necessarily correspond to what has happened *in the past*. While assessing the effects of policies in the past may have some utility – e.g., for electoral purposes – for a decisionmaker considering implementing a new policy or repealing a policy that is currently on the books, these backward-looking estimands are not germane without further assumptions.

## 3 Estimands

We require more policy-relevant estimands that are future-looking and correspond to "future causal effects." These estimands are designed to address actual policy decisions: should we implement or repeal this policy? We define them for a generic horizon $k \geq 1$ (i.e., $k$ periods ahead); conditioning on treatment status is at the end of the study period $p$ (e.g., effect for current adopters or non-adopters at horizon $k$).

Let

$$\tau_{i,p+k} = Y_{i,p+k}(1) - Y_{i,p+k}(0)$$

be the future individual treatment effect (FITE) for unit $i$ at horizon $k$. The FITE is arguably the most relevant quantity for

a citizen or policymaker belonging to unit $i$. For example, state lawmakers in California may likely be interested most in the FITE for California. The fact that individual treatment effects are not as a rule identifiable does not change the fact that it may be what is of primary importance. Do note also that none of the future causal effects we will discuss in this section are identifiable from the observed data without further assumptions.

Given the difficulties with estimating individual treatment effects, we may consider an alternative, the future average treatment effect among similar units (FATS). Given a distance measure $\rho$, we can define:

$$\tau_{i,p+k}(\eta;\rho) = \mathbb{E}_{\mathbb{P}_{p+k}}\left\{Y_{j,p+k}(1) - Y_{j,p+k}(0)|\rho(\mathbf{X}_i,\mathbf{X}_j) \leq \eta\right\}$$

i.e., the average treatment effect at $p+k$ among units $j$ such that $\rho(\mathbf{X}_i,\mathbf{X}_j) \leq \eta$; similarity is defined by distance in covariate space. (The backward-looking analog is the ATS: the average effect among units in that same set of units, under $\mathbb{P}_t$.) For example, policymakers in Massachusetts may be interested in the treatment effect among states in New England, or New York policymakers may be interested in the effect among other high-population states.

When policymakers do not have a unit-specific focus, other treatment effects may be of interest. Consider the future-looking analogs of typical backward-looking estimands at horizon $k$:

$$\tau_{p+k} = \mathbb{E}_{\mathbb{P}_{p+k}}\{Y_{i,p+k}(1) - Y_{i,p+k}(0)\} \quad\quad\quad (FATE)$$

$$\theta_{p+k} = \mathbb{E}_{\mathbb{P}_{p+k}}\{Y_{i,p+k}(1) - Y_{i,p+k}(0)|A_{ip} = 1\} \quad\quad\quad (FATT)$$

$$\omega_{p+k} = \mathbb{E}_{\mathbb{P}_{p+k}}\{Y_{i,p+k}(1) - Y_{i,p+k}(0)|A_{ip} = 0\} \quad\quad\quad (FATU)$$

The FATE encodes the average outcome at time $p+k$ if all units implemented or continued the policy compared to if all units discontinued or remained without the policy. The FATE is similar to the average treatment effect (ATE), and considers the most extreme counterfactual scenario of all implementing units versus none. The FATT encodes the difference in average outcome for units with the policy implemented at time $p$ (the end of the study period) if they continued the policy compared to if they discontinued the policy at time $p+k$. The FATT is analogous to the ATT and is targeted at the effect of repealing the policy. The FATU similarly encodes the difference in average outcome among units without the policy at time $p$ if they newly implemented the policy compared to if they remained without at time $p+k$. The FATU is most relevant to policymakers considering whether they should implement the policy.

Unlike the ATT and sub-ATTs above, none of the counterfactuals are observed. In fact, unlike for any backward-looking estimand, we lack any direct knowledge of the distribution of $\mathbb{P}_{p+k}$. We therefore require additional machinery to connect the observed data to the policy-relevant distribution $\mathbb{P}_{p+k}$. We outline below two different options: either taking time heterogeneity *less* seriously – by making time homogeneity assumptions – or taking it *more* seriously and formally modeling how policy effects change. **In the remainder of the paper we take $k = 1$ (one period ahead);** extension to $k > 1$ is analogous

and noted where it differs.

# 4   Taking dynamic effects less seriously: time homogeneity assumptions

There is such a rich and growing literature on estimation of the ATT and sub-ATTs [Wang et al., 2024, Sun and Abraham, 2021] that researchers may be interested in still using these tools to estimate forward-looking, policy-relevant estimands. Here, we outline in what cases this may be justified. We take as given that the researcher's design (e.g., parallel trends in DiD) has secured identification of the backward-looking $\theta$ or $\theta_{gt}$; the assumptions below are *extrapolation* assumptions that connect those quantities to the FATT.

## 4.1   Connecting the FATT to the ATT

We now consider what assumptions would be required to identify the FATT by the ATT. First, we could require homogeneity in effects that would rule out dynamic treatment effects:

**Assumption 1** (Strict ATT time homogeneity). $\mathbb{E}_{\mathbb{P}_t}\{Y_{it}(1) - Y_{it}(0)|A_{it} = 1\} = \mathbb{E}_{\mathbb{P}_s}\{Y_{is}(1) - Y_{is}(0)|A_{is} = 1\}$, *for all $s, t$.*

This assumption states that ATTs are time-invariant: the ATT at time $t$ is the same as at time $s$ for all $s, t$. We further need to place moderate limits on the between-state heterogeneity:

**Assumption 2** (Limited between-state ATT heterogeneity). $\mathbb{E}_{\mathbb{P}_{p+1}}\{Y_{i,p+1}(1) - Y_{i,p+1}(0)|A_{ip} = 1\} = \mathbb{E}_{\mathbb{P}_{p+1}}\{Y_{i,p+1}(1) - Y_{i,p+1}(0)|A_{i,p+1} = 1\}$

Assumption 2 limits the between-state heterogeneity such that the treatment effect among states who had implemented the law at the end of the study period (at time $p$) is the same as the treatment effect among states who will have implemented the law in the future (at time $p + 1$). In most cases, these two groups of states are nearly identical: most states that implemented the law at time $p$ keep it through to $p + 1$, and most states do not newly implement the law at time $p + 1$. So this assumption is generally relatively weak.

Under Assumptions 1 and 2, we have

$$\theta = \theta_{p+1}.$$

**Proposition 4.1** (Identification of the FATT under time homogeneity). *Suppose the backward-looking ATT $\theta$ is identified from the observed data under the researcher's design (e.g., parallel trends in difference-in-differences). Under Assumptions 1 and 2, the FATT $\theta_{p+1}$ is identified by $\theta_{p+1} = \theta$.*

**Remark 4.2.** *Because of the symmetry of their definitions, similar assumptions – with $A_{it} = 0$ replacing $A_{it} = 1$ and other similar replacements – would yield identification of the FATU with the ATU $\omega = \mathbb{E}_{\mathbb{P}}\{Y_{it}(1) - Y_{it}(0)|A_{it} = 0\}$.*

**Remark 4.3.** *Assumption 1 has testable implications. For any $s, t$ for which $\mathbb{E}_{\mathbb{P}_s}\{Y_{is}(1) - Y_{is}(0)|A_{is} = 1\}$ and $\mathbb{E}_{\mathbb{P}_t}\{Y_{it}(1) - Y_{it}(0)|A_{it} = 1\}$ are identifiable from the observable data, it may be tested whether they are equal. If differences between different time-specific treatment effects are small, that would lend credence to Assumption 1. In contrast, Assumption 2 is untestable: it concerns $\mathbb{P}_{p+1}$, for which we have no direct information in the observed data. Identification of the FATT under this path therefore rests on an extrapolation from period $p$ to $p + 1$—a form of transportability across time [Pearl and Bareinboim, 2022, Egami and Hartman, 2023]—and can only be supported by substantive reasoning or indirect checks. Similar to the assessment of parallel trends in difference-in-differences [Sun and Abraham, 2021, Wang et al., 2024], we may examine whether there are any large differences between $\mathbb{E}_{\mathbb{P}_{t+1}}\{Y_{i,t+1}(1) - Y_{i,t+1}(0)|A_{it} = 1\}$ and $\mathbb{E}_{\mathbb{P}_{t+1}}\{Y_{i,t+1}(1) - Y_{i,t+1}(0)|A_{i,t+1} = 1\}$ for any $t \in \{0, ..., p-1\}$ for which these effects are identified, as indirect evidence about stability of the composition of the treated group over time.*

## 4.2 Connecting other forward-looking estimands to backward-looking estimands

If we are willing to further restrict between-state heterogeneity, then we can find equivalence between other effects and the ATT. Consider assuming that there is no between-state heterogeneity in treatment effects at time $p + 1$:

**Assumption 3** (No future between-state ITE heterogeneity). *$\tau_{i,p+1} = \tau_{j,p+1}$ for all $i, j$.*

Under Assumptions 1 and 3, we have for all $i, \rho, \eta$,

$$\theta = \omega_{p+1} = \tau_{p+1} = \tau_{i,p+1}(\eta; \rho) = \tau_{i,p+1}.$$

Under these (restrictive) assumptions, all of the proposed forward-looking estimands are equivalent to the backward-looking ATT.

Often this assumption of strict ITE homogeneity may not be believable. While this may preclude using methods for estimating the ATT to identify these future-looking estimands, it does not preclude using corresponding backward-looking versions of these estimands under weaker assumptions. For example, under assumptions analogous to Assumptions 1 and 2, one has that $\tau_{p+1} = \tau \equiv \mathbb{E}_{\mathbb{P}}\{Y_{it}(1) - Y_{it}(0)\}$ by removing the conditioning on $A_{it} = 1$ from Assumptions 1 and 2. We state this in some generality in the following assumptions.

**Assumption 4** (Strict time homogeneity under $\mathscr{C}_t$). *$\mathbb{E}_{\mathbb{P}_t}\{Y_{it}(1) - Y_{it}(0)|\mathscr{C}_t\} = \mathbb{E}_{\mathbb{P}_s}\{Y_{is}(1) - Y_{is}(0)|\mathscr{C}_s\}$, for some conditioning event $\mathscr{C}_t$ defined on $\mathbb{P}_t$ and all $s, t \in \{1, ..., p+1\}$.*

**Assumption 5** (Limited between-state heterogeneity under $\mathscr{C}_t$). *$\mathbb{E}_{\mathbb{P}_{p+1}}\{Y_{i,p+1}(1) - Y_{i,p+1}(0)|\mathscr{C}_p\} = \mathbb{E}_{\mathbb{P}_{p+1}}\{Y_{i,p+1}(1) - Y_{i,p+1}(0)|\mathscr{C}_{p+1}\}$*

Under Assumptions 4 and 5, taking $\mathscr{C}_t = \{A_{it} = 0\}$ yields equivalence between the backward-looking ATU and the FATU; taking $\mathscr{C}_t = \emptyset$ yields equivalence between the backward-looking ATE and the FATE; and taking $\mathscr{C}_t = \{\rho(\mathbf{X}_i, \mathbf{X}_j) \leq \eta\}$ yields equivalence between the backward-looking ATS and the FATS (Assumption 5 not required for the FATS).

## 4.3 Connecting the FATT to sub-ATTs

Assumption 1 (and Assumption 4) restricts treatment effects to be constant over time. A great deal of recent literature in causal inference for panel data has explicitly rejected this assumption, positing that treatment effects may vary over time in multiple ways [Sun and Abraham, 2021, Brown and Butts, 2025]. Proposals for group-time effects $\theta_{gt}$ have attempted to capture variation in treatment effects by when the treatment was first implemented (by encoding this information in $G_i = g$) and also allowing treatment effects to vary dynamically over time thereafter.

One way to weaken Assumption 1 is to only require homogeneity within groups:

**Assumption 6** (Strict withing-group ATT time homogeneity). $\theta_{gt} = \theta_{gs} \equiv \theta_{g\cdot}$, for all $s, t \in \{1, ..., p+1\}$.

**Assumption 7** (Limited within-group between-state ATT heterogeneity). $\mathbb{E}_{\mathbb{P}_{p+1}}\{Y_{i,p+1}(1) - Y_{i,p+1}(0)|A_{ip} = 1, G_i = g\} = \mathbb{E}_{\mathbb{P}_{p+1}}\{Y_{i,p+1}(1) - Y_{i,p+1}(0)|A_{i,p+1} = 1, G_i = g\}$

Under Assumptions 6 and 7, we have that

$$\theta_{p+1} = \sum_g \mathbb{P}(G_i = g|A_{ip} = 1)\theta_{g\cdot}.$$

We assume that $G_i$ is a function of $A_i$ (e.g., time of first adoption), so that $\mathbb{P}(G_i = g|A_{ip} = 1)$ is the distribution of cohorts among units treated by period $p$. This states that, if within-group ATTs are not dynamic – i.e., they do not change over time – then a convex combination of group-specific ATTs may be used to extrapolate forward in time to time $p + 1$.

# 5 Taking dynamic effects more seriously: formal modeling

Much of the recent panel data literature with staggered adoption has either explicitly [Sun and Abraham, 2021, Wang et al., 2024, Brown and Butts, 2025] or implicitly [Stock and Watson, 2018] taken the view that assuming time homogeneity is too restrictive. Acknowledgment of the perverse effects of time heterogeneity in two-way fixed effects models [Sun and Abraham, 2021, Wang et al., 2024] spurred much of the recent literature on staggered adoption in difference-in-differences. However, unrestricted time heterogeneity renders future-looking estimands like the FATT inestimable and undercuts policy relevance of any estimated quantities. Instead, the heterogeneity of treatment effects must be restricted so that information from the observed data may be used to extrapolate to the future time period [Khosrowi, 2019, 2023, Maeba, 2024]. An alternative would be to report bounds or conduct sensitivity analysis without imposing a functional form [Manski, 2013]; the parametric path we take here trades that generality for point identification and extrapolation. Researchers should prefer this section's approach (Path 2) over the time-homogeneity path in Section 4 when they are unwilling to assume that within-group effects are constant over time but are willing to specify how effects evolve (e.g., as a function of event time).

To this end, to estimate the FATT, researchers must parameterize the time-varying ATTs or sub-ATTs in terms of time [Brown and Butts, 2025, Maeba, 2024]. For example, one could specify

$$\theta_{gt} = f(g, t; \boldsymbol{\gamma}), \qquad t = 1, ..., p+1; g \in \mathscr{G}$$

where $f(\cdot; \boldsymbol{\gamma})$ is a known function of a finite-dimensional parameter $\boldsymbol{\gamma} \in \mathbb{R}^d$. The parameter $\boldsymbol{\gamma}$ is identified from the observed group-time ATTs $\theta_{gt}$ for $t \leq p$, $g \in \mathscr{G}$, whenever the mapping from $\boldsymbol{\gamma}$ to $(\theta_{gt})_{g,t}$ (restricted to $t \leq p$) is injective, so that $\boldsymbol{\gamma}$ is uniquely determined by fitting the model to the identified $\theta_{gt}$. A common choice is to model effects as a function of event time $k = t - g$, e.g., $\theta_{gt} = h_g(k; \boldsymbol{\gamma})$ with $h_g$ linear or quadratic in $k$; then $f(g, t; \boldsymbol{\gamma}) = h_g(t - g; \boldsymbol{\gamma})$. Then the FATT may be identified as

$$\theta_{p+1} = \sum_g \mathbb{P}(G_i = g | A_{ip} = 1) f(g, p+1; \boldsymbol{\gamma}).$$

**Proposition 5.1** (Identification of the FATT under parametric dynamics). *Suppose the group-time ATTs $\theta_{gt}$ are identified from the observed data under the researcher's design for $t \leq p$, $g \in \mathscr{G}$. Suppose $\theta_{gt} = f(g, t; \boldsymbol{\gamma})$ with $\boldsymbol{\gamma}$ uniquely determined by $(\theta_{gt})_{g,t, t \leq p}$. Then the FATT $\theta_{p+1}$ is identified by $\theta_{p+1} = \sum_g \mathbb{P}(G_i = g | A_{ip} = 1) f(g, p+1; \boldsymbol{\gamma})$.*

## 5.1 Semiparametric estimation and EIF

We do not commit to a specific identification or first-stage estimation strategy. We assume that the group-time ATTs $\theta_{gt}$ are identified under the researcher's chosen design (e.g., difference-in-differences, synthetic control, or other methods); we do not restrict the identification strategy. We further assume that for each $(g, t)$, the researcher has access to an estimator $\widehat{\theta}_{gt}$ that is asymptotically linear with influence function $\phi_{gt}$: $\sqrt{n}(\widehat{\theta}_{gt} - \theta_{gt}) = n^{-1/2} \sum_{i=1}^n \phi_{gt,i} + o_{\mathbb{P}}(1)$. Such estimators exist in the literature for various designs [Callaway and Sant'Anna, 2021] and may be constructed with or without cross-fitting; our results hold for any such first-stage estimators.

The contribution of this subsection is to derive the efficient influence function (EIF) for the policy-relevant FATT $\theta_{p+1}$ as a functional of the group-time ATTs and their influence functions, so that any asymptotically linear estimators of $\theta_{gt}$ yield an asymptotically linear estimator of $\theta_{p+1}$ with limiting variance given by the propagated EIF.

**Path 1 (time homogeneity).** Under within-group time homogeneity, $\theta_{p+1} = \sum_g \omega_g \theta_{g\cdot}$ with $\omega_g = \mathbb{P}(G_i = g | A_{ip} = 1)$. Take as given the EIF $\phi_{\theta_{g\cdot}}$ for each $\theta_{g\cdot}$ (e.g., averaged over $t$ in the observed window, or as provided by the first-stage). By linearity, the EIF for $\psi_1 = \theta_{p+1}$ is

$$\phi_{\psi_1} = \sum_g \omega_g \phi_{\theta_{g\cdot}} + \sum_g \theta_{g\cdot} \phi_{\omega_g},$$

where $\phi_{\omega_g}$ is the influence function of the estimator of $\omega_g$ (e.g., for the empirical proportion among units with $A_{ip} = 1$, unit $i$ contributes $\mathbf{1}\{G_i = g, A_{ip} = 1\}/\mathbb{P}(A_{ip} = 1) - \omega_g)$; if $\omega_g$ is known, the second sum vanishes. Thus any asymptotically linear estimators of $\theta_{g.}$ yield an asymptotically linear estimator of $\psi_1$ with the stated influence function.

**Path 2 (extrapolation).** Under $\theta_{gt} = f(g, t; \boldsymbol{\gamma})$, we have $\theta_{p+1} = \Psi(\theta, \omega) = \sum_g \omega_g f(g, p+1; \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a smooth functional of $\theta = (\theta_{g,t})_{g,t}$. Take as given the EIF $\phi_{gt}$ for each $\theta_{gt}$ from whatever first-stage the researcher uses. By the chain rule, the EIF for $\psi_2 = \theta_{p+1}$ is

$$\phi_{\psi_2} = \frac{\partial \Psi}{\partial \theta} \phi_\theta + \frac{\partial \Psi}{\partial \omega} \phi_\omega,$$

with the obvious vectorization over $(g, t)$ and, if $\omega_g$ is estimated, an additional term $\sum_g f(g, p+1; \boldsymbol{\gamma})\phi_{\omega_g}$. The Jacobian $\frac{\partial \Psi}{\partial \theta}$ is the row vector with entries $\frac{\partial \Psi}{\partial \theta_{gt}} = \frac{\partial \Psi}{\partial \boldsymbol{\gamma}}^\top \frac{\partial \boldsymbol{\gamma}}{\partial \theta_{gt}}$, where $\frac{\partial \Psi}{\partial \boldsymbol{\gamma}} = \sum_g \omega_g \frac{\partial f(g, p+1; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}$; thus stacking the $\phi_{gt}$ and multiplying by this row gives the contribution from the first-stage. Within the model that restricts $\theta_{gt} = f(g, t; \boldsymbol{\gamma})$, this $\phi$ is the efficient influence function for $\theta_{p+1}$.

**Proposition 5.2** (Asymptotic distribution of $\widehat{\theta}_{p+1}$). *Under the assumptions of Path 1 or Path 2 above, suppose the first-stage estimators $\widehat{\theta}_{gt}$ (and, where relevant, $\widehat{\omega}_g$, $\widehat{\boldsymbol{\gamma}}$) are asymptotically linear with the influence functions stated. Then the estimator $\widehat{\psi} = \widehat{\theta}_{p+1}$ constructed by plugging these into the corresponding $\phi_{\psi_1}$ or $\phi_{\psi_2}$ satisfies $\sqrt{n}(\widehat{\psi} - \psi) \rightsquigarrow N(0, \sigma^2)$ with $\sigma^2 = \mathbb{E}[\phi^2]$.*

**Variance and asymptotics.** The semiparametric variance bound is $\sigma^2 = \mathbb{E}[\phi^2]$. Under standard regularity conditions, $\sqrt{n}(\widehat{\psi} - \psi) = n^{-1/2} \sum_{i=1}^n \phi_i + o_{\mathbb{P}}(1)$, so $\sqrt{n}(\widehat{\psi} - \psi) \rightsquigarrow N(0, \sigma^2)$. The variance may be estimated by $\widehat{\sigma}^2/n$ with $\widehat{\sigma}^2 = n^{-1} \sum_i \widehat{\phi}_i^2$ (or the sample variance of $\widehat{\phi}_i$); Wald confidence intervals are $\widehat{\psi} \pm z_{\alpha/2}\widehat{\sigma}/\sqrt{n}$. The estimated $\widehat{\phi}_i$ is a plug-in that uses $\widehat{\theta}_{gt}, \widehat{\boldsymbol{\gamma}}$, and possibly $\widehat{\omega}_g$; under standard regularity, this does not affect the first-order limit distribution. If the researcher uses flexible (e.g., machine-learning) estimation when forming $\widehat{\theta}_{p+1}$ (e.g., when estimating $\boldsymbol{\gamma}$ or $\omega_g$), $K$-fold cross-fitting can relax Donsker requirements: the aggregation or extrapolation step (estimate $\boldsymbol{\gamma}$ or $\omega_g$ on folds $\mathscr{I}_{-k}$, evaluate the EIF on $\mathscr{I}_k$) is cross-fit; only rate conditions (e.g., $o_{\mathbb{P}}(n^{-1/4})$) are needed. The paper does not require cross-fitting for the first-stage $\widehat{\theta}_{gt}$; the results hold for any asymptotically linear first-stage. A limitation of the parametric extrapolation approach is that bias and inference can be sensitive to the choice of $f$: if the temporal model is misspecified, $\widehat{\theta}_{p+1}$ need not be consistent for the FATT. In practice, researchers may compare several specifications (e.g., linear vs quadratic in event time) or report sensitivity of the extrapolated estimate to the choice of $f$; formal treatment of post-selection inference or misspecification is beyond this paper's scope.

# 6 Simulations

Policy-relevant questions concern the FATT at $p + 1$, whereas standard panel-data methods target backward-looking objects such as the ATT or group-time effects over the observed window. We use simulations to show when and how the two paths—time homogeneity (Path 1) and parametric extrapolation (Path 2)—succeed or fail to identify the FATT, and that EIF-based inference is valid when the model is correct. All designs use $q = 3$ groups, $p = 5$ periods, $n = 500$ units, and 1000 replicates.

## 6.1 Data generation

We consider $q$ cohorts (groups) and calendar times $1, \ldots, p$. For each group $g$, we observe group-time effects $\theta_{gt}$ for $t \geq g$ (post-treatment only), with event time $k = t - g$. The true effects $\theta_{gt}$ are generated from one of three specifications in event time: (i) constant within group, $\theta_{gt} = \theta_{g\cdot}$; (ii) linear, $\theta_{gt} = \alpha_g + \beta_g k$; (iii) quadratic, $\theta_{gt} = \alpha_g + \beta_g k + \delta_g k^2$. The strength of dynamics is thus controlled by the slope (and curvature) in event time. We take the role of a first-stage that has produced estimates $\widehat{\theta}_{gt}$ and their influence functions. For simplicity we set $\widehat{\theta}_{gt} = \theta_{gt} + \varepsilon_{gt}$ with $\varepsilon_{gt} \sim N(0, \sigma_\tau^2)$ and construct EIF vectors so that the variance of each cell estimator matches $\sigma_\tau^2$; aggregation and extrapolation then use the EIF propagation in Section 5. The simulations therefore focus on the aggregation and extrapolation step rather than on a specific difference-in-differences or other first-stage. The true FATT at $p + 1$ is $\theta_{p+1} = \sum_g \omega_g \theta_{g,p+1}$; we use equal group weights $\omega_g = 1/q$. The true backward-looking ATT is the (equal-weight) average of $\theta_{gt}$ over observed $(g, t)$.

The first simulation illustrates that the *target* of estimation matters: the same data can support an unbiased estimate of the backward-looking ATT while that number diverges from the FATT when effects evolve with event time. When the slope in event time is non-zero, the two curves in Figure 1 separate; interpreting the backward-looking ATT as the policy-relevant effect is then misleading. This motivates the need for either time-homogeneity assumptions (Path 1) or explicit extrapolation (Path 2).

Path 1 assumes within-group time homogeneity and forms $\widehat{\theta}_{p+1} = \sum_g \omega_g \bar{\theta}_{g\cdot}$, where $\bar{\theta}_{g\cdot}$ is the average of $\widehat{\theta}_{gt}$ over $t$ for group $g$. Scenario A (constant effects within group) shows small bias and coverage near 95%, consistent with Section 4.

Table 1: Path 1: time homogeneity (Scenario A) vs dynamics (Scenario B).

| Scenario | True FATT | Bias | RMSE | 95% coverage |
|---|---|---|---|---|
| A (homogeneity) | 0.133 | 0.001 | 0.031 | 94.10 % |
| B (dynamics) | 0.537 | -0.313 | 0.315 | 0.00 % |

Scenario B (linear dynamics) shows substantial bias and coverage collapsing to zero for the FATT—Path 1 then estimates a different object (the constant-within-group extrapolant) and is invalid for the FATT. This underscores that Path 1 is only appropriate when time homogeneity is plausible.

Path 2 fits a parametric model $f(g, t; \widehat{\gamma})$ to $\widehat{\theta}_{gt}$ and extrapolates to $p + 1$. Correct specification (linear DGP, linear fit)
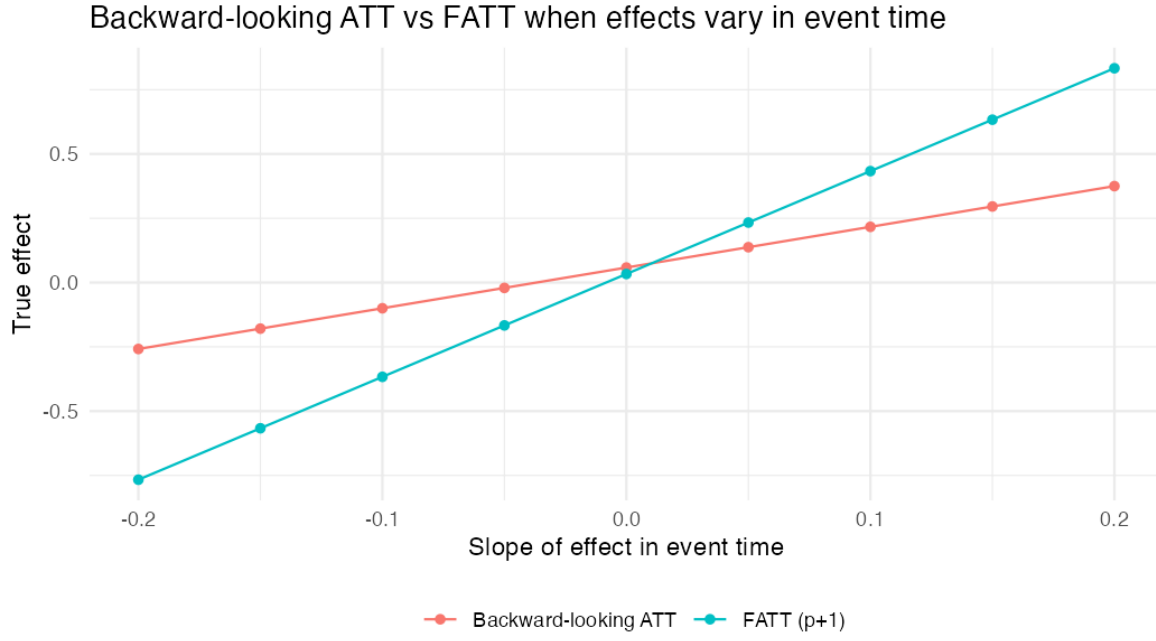
Figure 1: True backward-looking ATT and true FATT at $p+1$ as a function of the slope of the effect in event time (linear DGP). When the slope is non-zero, the two estimands differ; interpreting the backward-looking ATT as the policy-relevant effect is misleading.

Table 2: Path 2: correct specification (linear DGP, linear fit), misspecified (quadratic DGP, linear fit), and quadratic fit (quadratic DGP, quadratic fit).

| Specification | True FATT | Bias | RMSE | 95% coverage |
|---|---|---|---|---|
| Correct (linear fit) | 0.427 | -0.001 | 0.074 | 95.70 % |
| Misspecified (linear on quad) | 0.710 | -0.147 | 0.166 | 49.70 % |
| Quadratic fit on quad | 0.710 | 0.008 | 0.189 | 94.30 % |

yields low bias and nominal coverage. Misspecification (quadratic DGP, linear fit) introduces bias and can distort coverage; fitting the correct quadratic model restores performance. This illustrates that Path 2 trades robustness for point identification and that the choice of $f$ matters in practice.

Under correct specification, the EIF-based variance and Wald intervals from Section 5 should achieve nominal coverage. The variance ratio near one and coverage near 95% and 90% confirm that the propagated EIF delivers valid inference when

Table 3: EIF-based variance and Wald coverage (Path 2, correct specification).

| Metric | Value |
|---|---|
| Variance ratio (est./emp.) | 0.95 |
| 95% coverage | 94.90 % |
| 90% coverage | 89.00 % |

the first-stage and (for Path 2) the extrapolation model are correct.

On a single DGP with mild dynamics, we compare both paths. Path 2 reduces bias for the FATT relative to Path 1

Table 4: Path 1 vs Path 2 on the same DGP with mild dynamics (target: true FATT).

| Estimator | Bias | RMSE | 95% coverage |
|---|---|---|---|
| Path 1 | -0.149 | 0.152 | 0.00 % |
| Path 2 | 0.003 | 0.071 | 95.40 % |

when effects evolve with event time; Path 1's simplicity is attractive when dynamics are absent, but when they are present, extrapolation can improve alignment with the policy-relevant FATT.

Taken together, the simulations reinforce the paper's message: the target of estimation (FATT vs backward-looking ATT) matters; Path 1 and Path 2 each have their place depending on whether the researcher is willing to assume time homogeneity or to model dynamics; and EIF-based inference is well-calibrated under correct specification.

# 7  Discussion

We have argued that policy analysis with panel data faces a tension: methods that allow dynamic, group-time effects address important identification concerns but produce backward-looking estimands that do not directly answer whether to maintain or adopt a policy now or next period. We resolve this by defining future causal effects (FATT, FATU, FATE, FATS) and offering two identification paths. **Path 1 (time homogeneity):** Under strict within-group time homogeneity and limited between-state heterogeneity, the FATT is identified by the backward-looking ATT or by a convex combination of group-specific ATTs; the researcher can use standard group-time or ATT estimators and aggregate. **Path 2 (extrapolation):** When effects are allowed to vary with event time (or calendar time), the FATT is identified by fitting a parametric model $f(g, t; \gamma)$ to the identified group-time ATTs and extrapolating to $p + 1$. The choice between paths is substantive: Path 1 is appropriate when stability

of effects over time is plausible (and can be partly checked); Path 2 is appropriate when the researcher is willing to model dynamics and to accept sensitivity to the choice of $f$.

Several limitations deserve emphasis. Identification of the FATT in both paths relies on the distribution $\mathbb{P}_{p+1}$ (and, for Path 2, on the extrapolation model holding at $p+1$), which is not directly observable. We have focused on a single future period ($k = 1$); extension to multiple horizons $k > 1$ is conceptually straightforward but adds assumptions on how far the model can be projected. Inference uses the propagated efficient influence function, which is valid under correct specification of the first-stage and, for Path 2, of $f$; misspecification can bias the extrapolated estimate and distort inference.

Natural next steps include sensitivity analysis for violations of time homogeneity or of the extrapolation model, formal treatment of model selection or multiple specifications, and applications that compare Path 1 and Path 2 on the same data and discuss which assumptions are more plausible in context.

When researchers are unwilling to impose a parametric $f$, an alternative to point-identifying the FATT is to forego point identification and report bounds under weak assumptions—for example, that future effects lie in a bounded or monotone class—as in the partial-identification and sensitivity literature [Manski, 2013]. That approach yields an interval that contains the FATT rather than a single number. Path 2 can be interpreted as assuming that the fitted parameter $\gamma$ encodes structure that is invariant to the time period, so that extrapolation to $p+1$ is valid so long as this structure holds; this connects to the literature on policy-invariant parameters and transportability [Heckman and Vytlacil, 2005, Pearl and Bareinboim, 2022]. In practice, researchers can also compare several specifications of $f$ (e.g., linear vs quadratic in event time) and report how the extrapolated FATT and its interval change, as noted in Section 5.

# References

Dmitry Arkhangelsky and Guido Imbens. Causal models for longitudinal and panel data: A survey. *The Econometrics Journal*, 27(3):C1–C61, 2024.

Andrew Baker, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. Difference-in-Differences Designs: A Practitioner's Guide, June 2025.

Elias Bareinboim and Judea Pearl. Transportability of causal effects: Completeness results. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pages 698–704, 2012.

Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. Inferring causal impact using Bayesian structural time-series models. 2015.

Nicholas L. Brown and Kyle Butts. Dynamic treatment effect estimation with interactive fixed effects and short panels. *Journal of Econometrics*, 250:106013, 2025.

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225 (2):200–230, 2021.

Nancy Cartwright. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press, 2007.

Dakota W. Cintron, Nancy E. Adler, Laura M. Gottlieb, Erin Hagan, May Lynn Tan, David Vlahov, Madellena Maria Glymour, and Ellicott C. Matthay. Heterogeneous treatment effects in social policy studies: An assessment of contemporary articles in the health and social sciences. *Annals of epidemiology*, 70:79–88, June 2022. ISSN 1047-2797. doi: 10.1016/j.annepidem. 2022.04.009.

Stephen R. Cole and Elizabeth A. Stuart. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *American journal of epidemiology*, 172(1):107–115, 2010.

Clément De Chaisemartin and Xavier d'Haultfoeuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The econometrics journal*, 26(3):C1–C30, 2023.

Holger Dette and Martin Schumann. Testing for Equivalence of Pre-Trends in Difference-in-Differences Estimation. *Journal of Business & Economic Statistics*, 42(4):1289–1301, October 2024. ISSN 0735-0015, 1537-2707. doi: 10.1080/07350015. 2024.2308121.

Martin Devaux and Naoki Egami. Quantifying robustness to external validity bias. *Available at SSRN 4213753*, 2022.

Naoki Egami and Erin Hartman. Elements of external validity: Framework, design, and analysis. *American Political Science Review*, 117(3):1070–1088, 2023.

Neil R. Ericsson, David F. Hendry, and Grayham E. Mizon. Exogeneity, Cointegration, and Economic Policy Analysis. *Journal of Business & Economic Statistics*, 16(4):370–387, October 1998. ISSN 0735-0015, 1537-2707. doi: 10.1080/07350015. 1998.10524778.

Arturo Estrella and Jeffrey C. Fuhrer. Are" deep" parameters stable?: The lucas critique as an empirical hypothesis. Technical report, Federal Reserve Bank of Boston Boston, MA, 1999.

Laura Forastiere, Fan Li, and Michela Baccini. Forecasting Causal Effects of Future Interventions: Confounding and Transportability Issues, July 2025.

Simon Freyaldenhoven, Christian Hansen, and Jesse M. Shapiro. Pre-event trends in the panel event-study design. *American Economic Review*, 109(9):3307–3338, 2019.

Nicole B Gabler, Naihua Duan, Diana Liao, Joann G Elmore, Theodore G Ganiats, and Richard L Kravitz. Dealing with heterogeneity of treatment effects: Is the literature up to the challenge? *Trials*, 10:43, June 2009. ISSN 1745-6215. doi: 10.1186/1745-6215-10-43.

Christian Gische, Stephen G. West, and Manuel C. Voelkle. Forecasting Causal Effects of Interventions versus Predicting Future Outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(3):475–492, May 2021. ISSN 1070-5511, 1532-8007. doi: 10.1080/10705511.2020.1780598.

James J Heckman and Edward Vytlacil. Policy-Relevant Treatment Effects. *American Economic Review*, 91(2):107–111, May 2001. ISSN 0002-8282. doi: 10.1257/aer.91.2.107.

James J. Heckman and Edward Vytlacil. Structural Equations, Treatment Effects, and Econometric Policy Evaluation1. *Econometrica*, 73(3):669–738, May 2005. ISSN 0012-9682, 1468-0262. doi: 10.1111/j.1468-0262.2005.00594.x.

Donal Khosrowi. Extrapolation of causal effects – hopes, assumptions, and the extrapolator's circle. *Journal of Economic Methodology*, 26(1):45–58, January 2019. ISSN 1350-178X, 1469-9427. doi: 10.1080/1350178X.2018.1561078.

Donal Khosrowi. Extrapolating from experiments, confidently. *European Journal for Philosophy of Science*, 13(2):18, March 2023. ISSN 1879-4920. doi: 10.1007/s13194-023-00520-1.

Robert E. Lucas Jr. Econometric policy evaluation: A critique. In *Carnegie-Rochester Conference Series on Public Policy*, volume 1, pages 19–46. North-Holland, 1976.

Kensuke Maeba. Extrapolation of Treatment Effect Estimates Across Contexts and Policies: An Application to Cash Transfer Experiments, March 2024.

Charles F. Manski. *Public Policy in an Uncertain World: Analysis and Decisions*. Harvard University Press, 2013.

Judea Pearl and Elias Bareinboim. External Validity: From *Do* -Calculus to Transportability Across Populations. In Hector Geffner, Rina Dechter, and Joseph Y. Halpern, editors, *Probabilistic and Causal Inference*, pages 451–482. ACM, New York, NY, USA, 1 edition, February 2022. ISBN 978-1-4503-9586-1. doi: 10.1145/3501714.3501741.

Ashesh Rambachan and Jonathan Roth. A more credible approach to parallel trends. *Review of Economic Studies*, 90(5): 2555–2591, 2023.

Jonathan Roth. Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322, 2022.

Jonathan Roth and Pedro H. C. Sant'Anna. Efficient Estimation for Staggered Rollout Designs. *Journal of Political Economy Microeconomics*, 1(4):669–709, November 2023. ISSN 2832-9368, 2832-9376. doi: 10.1086/726581.

Yuya Sasaki and Takuya Ura. Estimation and inference for policy relevant treatment effects. *Journal of Econometrics*, 234(2): 394–450, 2023.

Ilya Shpitser. Counterfactual Graphical Models for Longitudinal Mediation Analysis With Unobserved Confounding. *Cognitive Science*, 37(6):1011–1035, August 2013. ISSN 0364-0213, 1551-6709. doi: 10.1111/cogs.12058.

James H. Stock and Mark W. Watson. Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610):917–948, 2018.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, December 2021. ISSN 0304-4076. doi: 10.1016/j.jeconom.2020.09.006.

Elizabeth Tipton. Improving Generalizations From Experiments Using Propensity Score Subclassification: Assumptions, Properties, and Contexts. *Journal of Educational and Behavioral Statistics*, 38(3):239–266, June 2013. ISSN 1076-9986, 1935-1054. doi: 10.3102/1076998612441947.

Guangyi Wang, Rita Hamad, and Justin S. White. Advances in Difference-in-differences Methods for Policy Evaluation Research. *Epidemiology (Cambridge, Mass.)*, 35(5):628–637, September 2024. ISSN 1044-3983. doi: 10.1097/EDE.0000000000001755.